On Improving the Quality of Quarterly GDP in JAPAN : A Recent Attempt

Naoto Kunitomo
Graduate School of Economics, University of Tokyo
Bunkyo-ku Hongo 7-3-1
Tokyo 113-0033, Japan
E-mail: kunitomo@e.u-tokyo.ac.jp

Seisho Sato
Institute of Statistical Mathmeatics
Midoricho 10-3
Tachikawa-shi 190-8562, Japan
E-mail: sato@ism.ac.jp

Shigeru Hirota
Cabinet Office, Government of Japan
Chiyoda-ku, Kasumigaseki 3-1
Tokyo 100-8970, Japan
E-mail: shigeru.hirota@cao.go.jp

Introduction

As the growth rate of the Japanese macro-economy has been slow-downed in the past fifteen years, the accuracy of the (official) quarterly GDP figures has become much more important than in the era of high-growth period for the official statistical agency within the central government of Japan, which has been in charge of Quarterly GDP. There have been critical comments on the accuracy of Japanese GDP from some economists and mass-media, and then the statistical agency has been trying to make the quarterly (quick) GDP more accurate. We have been proposing to use the modern statistical time series analysis including the state space modeling in particular to improve the quality of quarterly (quick) estimates of GDP in JAPAN. Some statistical methods we proposed have been adopted already in practice, but there would be further problems on the current estimation method to be improved.

Fixed Investments and Quarterly GDP

One of major problems to compute Quarterly Quick GDP Estimates in JAPAN has been the fact that the flow of new information available depends on the timing of various statistical surveys and many of them are conducted by different organizations (ministries) within the central government. When the Cabinet Office needs to publish the first quick estimate of Quarterly GDP, one important information of fixed investments called *Hojin-Kigyo-Tokei*, which is a large scale of survey on the private firms, conducted by the Ministry of Finance is not available while we have an estimate of the fixed investment from the supply side estimates (based on commodity-flow information). After the first estimate is published, the important source of information on fixed investments from the demand side becomes available and then the new number should be incorporated with the supply side information to form the 2nd quick Quarterly estimate (a revision of the 1st quick estimate). Hence at the initial state of estimating quarterly GDP, we have a missing observation problem and we need to predict the coming number of fixed investment in the demand side. Also we have similar problems on some components of inventory investments in the process of quick Quarterly GDP estimations. Since the macro-economic data have trends, cycles, seasonality and irregularity components, the predictions of the fixed investments and inventory investments are rather difficult. The missing observation and prediction problem we have are rather complexed and challenging in the proper statistical sense.

Modeling fixed investments

We take the (private) fixed investments problem as an illustration. Let the $j-$th observation of the fixed investments from the demand side be X_{dj} and the $j-$th observation of the fixed investments from the supply side be X_{sj}. We assume that X_{ij} $(i = d, s; j = 1, \cdots, N)$ satisfy

(1) $$X_{ij} = T_{ij} + C_{ij} + S_{ij} + I_{ij} \ ,$$

where T_{ij}, C_{ij}, S_{ij}, and I_{ij} are the trend components, the cycle components, the seasonal components and the irregular components, respectively. For the simplicity, we treat the sum of trend and cycle components as $TC_{ij} = T_{ij} + C_{ij}$, and then we have

(2) $$X_{ij} = TC_{ij} + S_{ij} + I_{ij} \ .$$

By denoting $X_{ij}^{(a)} = X_{ij} - S_{ij}$, $X_{ij}^{(a)} = TC_{ij} + I_{ij}$ and

(3) $$\begin{aligned} \Delta X_{ij}^{(a)} &= X_{ij}^{(a)} - X_{i,j-1}^{(a)} \\ &= (TC_{ij} - TC_{i,j-1}) + (I_{ij} - I_{i,j-1}) \\ &= \Delta TC_{ij} + \Delta I_{ij} \ , \end{aligned}$$

where ΔX_{ij} $(= X_{ij} - X_{i,j-1})$.

Because two components should have the same true fixed investments series, we assume $TC_{dj} = TC_{sj} = TC_j$ and then we have

(4) $$\Delta X_{dj}^{(a)} = \Delta TC_j + \Delta I_{dj} \ , \ \Delta X_{sj}^{(a)} = \Delta TC_j + \Delta I_{sj} \ .$$

In this formulation we need to consider a statistical prediction problem that given the information available at $j = n-1$ $\Delta X_{dj}^{(a)}, \Delta X_{sj}^{(a)}$ $(j = 1, \cdots, n-1)$ and $\Delta X_{sn}^{(a)}$, we need to predict $\Delta X_{dn}^{(a)}$. The first quick estimate of the private fixed investement is calculated as a linear combination of two estimates from different sources by the present estimation method. We denote the conditional expectation as the optimal predictor

(5) $$Y_n = \mathcal{E}[\Delta X_{dn}^{(a)} | I_{n-1}] \ ,$$

where $\mathcal{E}[\ \cdot\ | I_{n-1}]$ is the conditional expectation, I_{n-1} is the information at $n-1$ and $\Delta X_{sn}^{(a)} \in I_{n-1}$.

We note that the present formulation is quite similar to the official explanation of the Quarterly GDP manual in Japan (Cabinet Office (2006)), which could be interpreted as the errors-in-variables model or the measurement-errors model. Furthermore, we assume that for any j, k $(j, k = 1, \cdots, N)$ (A) ΔTC_j and I_{dk}, I_{sk} are independent, (B) ΔTC_j and I_{dk}, I_{sk} follow the normal distributions with the initial condition ΔTC_1. (It is a rather strong condition, but usually we do not have many observations for the proper use of the non-parametric methods.)

Then for $j = 2, \cdots, n$

(6) $$\begin{bmatrix} \Delta TC_j \\ I_{dj} \\ I_{sj} \end{bmatrix} \sim N\left(\begin{bmatrix} \xi_j \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_{\xi\xi} & 0 & 0 \\ 0 & \sigma_{dd} & \sigma_{ds} \\ 0 & \sigma_{sd} & \sigma_{ss} \end{bmatrix} \right) .$$

As the simplest situation we may assume that the expected value of ΔTC_j is constant ($\xi_j = \xi$), $\sigma_{\xi\xi}$ is the variance of ΔTC_i, and $\sigma_{dd}, \sigma_{ss}, \sigma_{ds}$ are the variances and covariance of the irregular components in the demand side and supply side, respectively.

Given the current information $\Delta X_{sn}^{(a)}$, the optimal predictor of $\Delta X_{dn}^{(a)}$ is given by

$$(7) \qquad \begin{aligned} \mathcal{E}[\Delta X_{dn}^{(a)}|\Delta X_{sn}^{(a)}] &= \mathcal{E}[\Delta X_{dn}^{(a)}] + \frac{Cov(\Delta X_{dn}^{(a)}, \Delta X_{sn}^{(a)})}{V(\Delta X_{sn}^{(a)})}\left(\Delta X_{sn}^{(a)} - \mathcal{E}[\Delta X_{sn}^{(a)}]\right) \\ &= \left(\frac{\sigma_{\xi\xi} + 2\sigma_{ds}}{\sigma_{\xi\xi} + 2\sigma_{ss}}\right)\Delta X_{sn}^{(a)} + \left(1 - \frac{\sigma_{\xi\xi} + 2\sigma_{ds}}{\sigma_{\xi\xi} + 2\sigma_{ss}}\right)\xi_n \ , \end{aligned}$$

where the coefficient of $\Delta X_{sn}^{(a)}$ is a regression coefficient and the predictor is a weighted average of $\Delta X_{sn}^{(a)}$ and ξ_n. Then MSE (the prediction mean-squared error) is

$$(8) \qquad \begin{aligned} MSE_1 &= \mathcal{E}\left[(Y_n - \Delta X_{dn}^{(a)})^2\right] \\ &= (\sigma_{\xi\xi} + 2\sigma_{dd})[1 - \frac{(\sigma_{\xi\xi} + 2\sigma_{ds})^2}{(\sigma_{\xi\xi} + 2\sigma_{ss})(\sigma_{\xi\xi} + 2\sigma_{dd})}] \\ &= [2\sigma_{ss} + 2\sigma_{dd} - 4\sigma_{ds}] - 4\frac{(\sigma_{ss} - \sigma_{ds})^2}{(\sigma_{\xi\xi} + 2\sigma_{ss})} \ . \end{aligned}$$

We note that we have the predictor as if the variances and covariance were known in advance, but in practice we need to estimate them and use ξ_n.

One possible predictor would be $\Delta X_{sn}^{(a)}$, which was used in the past, and its (prediction) MSE is given by

$$(9) \qquad \begin{aligned} MSE_1' &= \mathcal{E}[\Delta X_{sn}^{(a)} - \Delta X_{dn}^{(a)}]^2 \\ &= \mathcal{E}[\Delta I_{sn}^{(a)} - \Delta I_{dn}^{(a)}]^2 \\ &= 2\sigma_{ss} + 2\sigma_{dd} - 4\sigma_{ds} \ . \end{aligned}$$

Then we have found that it is not an optimal predictor and and MSE_1' cannot be less than MSE_1. In the errors-in-variables model σ_{ds} is not necessarily equal to σ_{ss} and the regression coefficient of $\Delta X_{sn}^{(a)}$ is not 1. From (8) and (9), as the variance of ΔTC, $\sigma_{\xi\xi} \to 0$, we could improve the prediction.

If we knew the current ΔTC_n as an extra information, we have the predictor of ΔX_{dn} is given by

$$(10) \qquad \begin{aligned} \mathcal{E}[\Delta X_{dn}^{(a)}|\Delta TC_n] &= \xi_n + \frac{Cov(\Delta X_{dn}^{(a)}, \Delta TC_n)}{V(\Delta TC_n)}(\Delta TC_n - \xi_n) \\ &= \xi_n + (\Delta TC_n - \xi_n) \\ &= \Delta TC_n \ . \end{aligned}$$

Then MSE is given by

$$(11) \qquad MSE_2 = (\sigma_{\xi\xi} + 2\sigma_{dd})[1 - \frac{(\sigma_{\xi\xi})^2}{\sigma_{\xi\xi}(\sigma_{\xi\xi} + 2\sigma_{dd})}] = 2\sigma_{dd} \ ,$$

which is $\mathcal{V}[\Delta I_{dn}]$. Hence

$$(12) \qquad MSE_1 - MSE_2 = \frac{2}{(\sigma_{\xi\xi} + 2\sigma_{ss})}[\sigma_{\xi\xi}\sigma_{s,s} - 2\sigma_{ds}(\sigma_{ds} + \sigma_{\xi\xi})] \ .$$

If σ_{ds} was small, then we have $MSE_1' > MSE_2, MSE_1 > MSE_2$. We have an intuition that we did not have the irregular component in the supply side, we have better prediction.

More generally, we consider the situation that we have information of not only ΔTC_n, but also past information ΔTC_j $(j \le n-1)$. Then we have the errors-in-variables model

$$(13) \qquad \begin{bmatrix} \Delta TC_j \\ \Delta TC_{j-1} \\ I_{dj} \\ I_{sj} \end{bmatrix} \sim N\left(\begin{bmatrix} \xi_j \\ \xi_{j-1} \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_{\xi_j,\xi_j} & \sigma_{\xi_j,\xi_{j-1}} & 0 & 0 \\ \sigma_{\xi_{j-1},\xi_j} & \sigma_{\xi_{j-1},\xi_{j-1}} & 0 & 0 \\ 0 & 0 & \sigma_{dd} & \sigma_{ds} \\ 0 & 0 & \sigma_{sd} & \sigma_{ss} \end{bmatrix}\right) ,$$

σ_{ξ_j,ξ_j} is the variance of ΔTC_j, and $\sigma_{\xi_j,\xi_{j-1}}$ is the covariance of ΔTC_j and ΔTC_{j-1}.

When the distributions of the underlying random variables are normal, the optimal predictor of ΔTC_n given ΔTC_j $(j \leq n)$ is a linear combination of ΔTC_j $(j \leq n)$. Under the errors-in-variables model (1), (2) AND (6)M we have the following result.

Theorem 1 : Assume the conditions (A) and (B) in (1), (2) and (6) with the initial condition. The optimal predictor with the minimum mean squares error is given by

$$\mathcal{E}[\Delta X_{dn}^{(a)} | \Delta TC_n, \Delta TC_j \ (j \leq n-1)] = \Delta TC_n \tag{14}$$

and the (prediction) mean squared error is $MSE_2 = 2\sigma_{dd}$.

Also we can consider the situation when we use the information on the irregular part in the supply side I_{sn}. However, since we have $\mathbf{Cov}[\Delta TC_n, \Delta I_{sn}] = 0$ and $\mathbf{Cov}[\Delta X_{dn}^{(a)}, \Delta I_{sn}] = 2\sigma_{ds}$, we summarize our arguments.

Corollary 1 : In the errors-in-the variables model (1), (2) and (6), in addition to the conditions (A) and (B), assume (C) $\sigma_{ds} = 0$. Then the minimum mean squared error predictor is given by

$$\mathcal{E}[\Delta X_{dn}^{(a)} | \Delta TC_n, I_{sn}, \Delta TC_j \ (j \leq n-1)] = \Delta TC_n \tag{15}$$

and the predictive MSE is $MSE_2 = 2\sigma_{dd}$.

We give several remarks on the present statistical problem.
(i) In comparing the predictor ΔX_{sn}, we have better predictors based on $\Delta TC_n, \Delta I_{sn}$ in the sense of MSE if the estimation of unobservable states is reasonable. By using the fixed investments data, we have found that the estimate of σ_{ds} is small (and not statistically significant) and then it would be unlikely to have better predictor by using ΔI_{sn}. It may be reasonable to expect to have some noise term in our predictors. When the covariance is zero, ΔTC_n is a sufficient statistics for $\Delta X_{dn}^{(a)}$.
(ii) Our arguments are based on the assumption that we can estimate the states of different components of time series (trends, ctcles, seasonality and irregular components) by using DECOMP or X-12-ARIMA. We have used the DECOMP program to estimate the components by using the state space models.
(iii) The problem of seasonal adjustments complicates the situation and we have used the DECOMP program by Kitagawa (2010) to estimate the seasonality components of the fixed investments.

Some results on fixed investments

We have used the fixed investments of 1994Q1∼2009Q4 in the demand side and supply side investments. We first extract the trend components, the cycle components, the seasonal components and the irregular components by using the DECOMP program. For an illustration, we show the decomposition result of an official (aggregate) fix investments as in Figure 1.

(i) In this period, the official time series of (private) fixed investments have been fluctuated rather wildly. We have found a clear seasonal pattern up to around 2006, but since then the movements of investments have been changed to some extent. However, it is quite difficult to make an change point analysis because we do not have enough series.
(ii) The correlation between the irregularity components in the demand side and supply side of the fixed investments is quite small, while the correlation between the trend components and cycle components is quite high. Therefore it may be reasonable to use the prediction method based on the supply side estimation of ΔTC_n.
(iii) We have tried to use other prediction methods based on $\Delta TC_{n-j}, j \geq 1$, but we could not have found significant improvements on the predictive power although the predictors become more complicated. We also have tried to use several non-linear prediction methods based on the non-linear

modeling of the fixed investment series, but we could not have found better results.

(iv) We have compared several prediction methods for the quick quarterly estimates (the 1st estimate, the second estimate) and their predictive powers on the subsequent final estimate. We have found that the predictor based on the estimate of the trend-cycle components of the fixed investments has reasonable results.

Concluding Remarks

We have discussed some important issues on the past and current estimation method of the (quick) quarterly GDP in Japan. Since economists and policy makers within the central government of Japan have been interested in the current state of macro-economy and the main macro time series has been the official (quick) GDP estimates, it is important to improve the statistical method of estimation. While we want to known the GDP figure as soon as possible, but at the same time we need to estimate the GDP based on less precise economic data. Since one of major source of inaccuracy on quick quarterly GDP has been the estimation of (private) fixed investments, we have proposed to use the errors-in-variables model (or the measurement error model) and the state space modeling for decomposing the trend-cycle components, the seasonal components and the irregular components. Then we have used the estimate of the trend-cycle components to form the quick quarterly fixed investments. Based on the Japanese GDP data, we have found some improvements on the quality of quick GDP estimation. (The details have been discussed in Sato and Kunitomo (2010).)

There have been a number of further problems to be dealt with. The current seasonal adjustments practice including the current X-12-ARIMA procedure in JAPAN should be examined in particular. There have been unexpected macro-shocks in the Japanese economy and then the current estimation method for quarterly GDP should be under further investigation.

REFERENCES

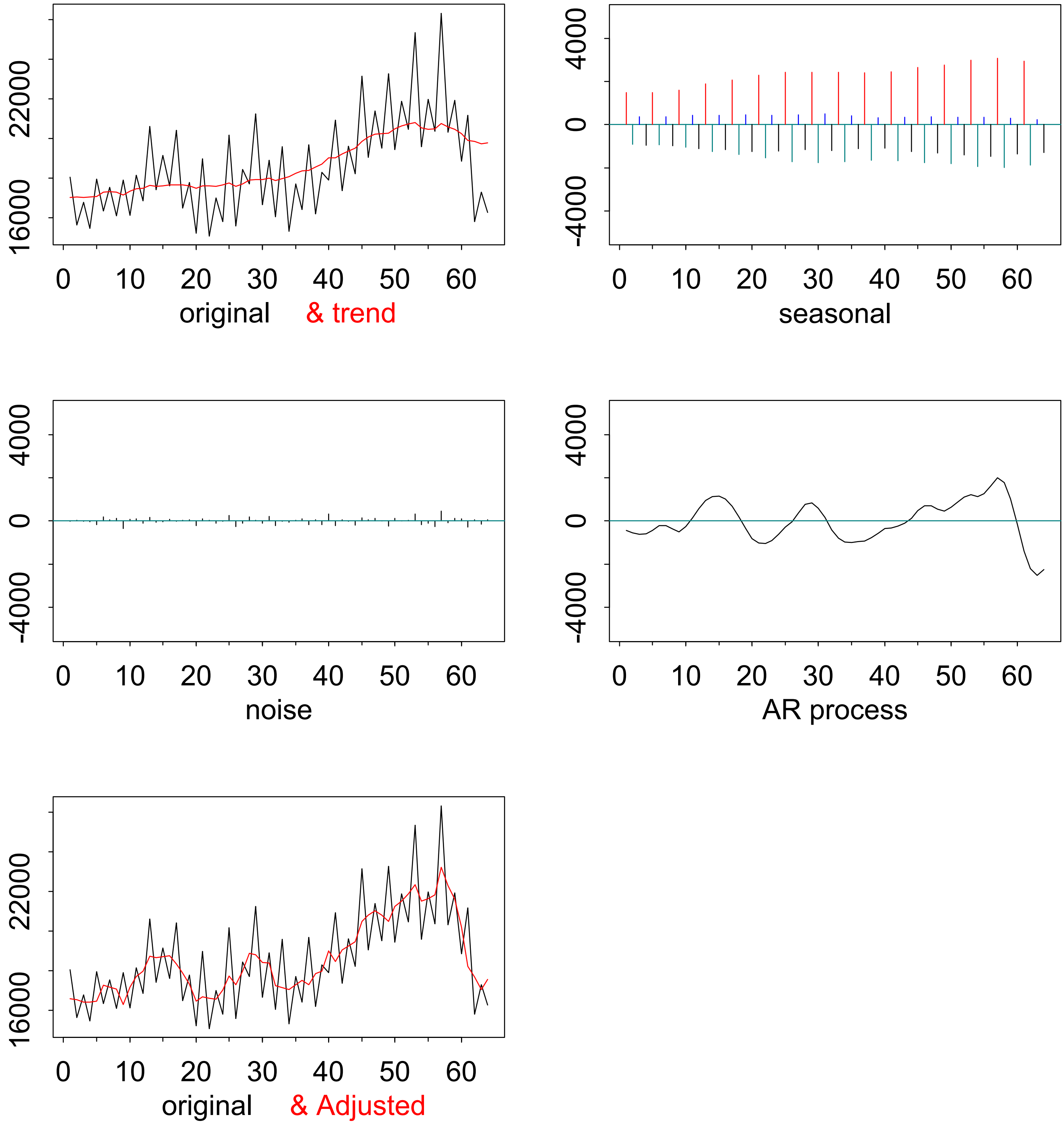
[1] IMF Quarterly GDP Estimation Manual (2001), International Monetary Fund (Home Page).

[2] Kitagawa G. (2010) *Introduction to Time Series Analysis*, CRC Press (Talor & Francis Group).

[3] Kunitomo, N. and S. Sato (2010) "On Improving Quarterly GDP in JAPAN," (in Japanese), Socio-economic Research Institute, Cabinet Office, Discussion Paper No.249 (http://www.ersi.go.jp/jp/archive/e_dis/e_discus.html/).

[4] Department of National Accounts, Cabinet Office (2006), "Estimation Method of Quarterly GDP," (http://www.ersi.cao.go.jp/jp/sna/sokuhou/gaiyou/ sakusei/sakusei_top.html/).

Figure 1 : A decomposition of Fixed Investments


16000
22000
0
10
20
30
40
50
60
original & trend
-4000
0
4000
seasonal
noise
AR process
original & Adjusted